

Fair Digital Object Profile for Marine Images

The iFDO Standard, Ecosystem and Life Cycle

Karsten Schimpf ¹, Christopher Krämmer ¹, Timm Schoening ²

¹Alfred Wegener Institute for Polar and Marine Research, Bremerhaven, Germany

²GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany

*Correspondence: Karsten Schimpf, karsten.schimpf@awi.de

Abstract. The heterogeneity inherent in the content and context of marine images poses a significant challenge for FAIR (Findable, Accessible, Interoperable, and Reusable) publishing principles, as different scientific disciplines possess distinct standards and terminologies that impede cross-domain data reuse. The development of a common FAIR Digital Object (FDO) profile tailored for marine images, referred to as iFDO, has the potential to dismantle these obstacles by instituting a standardized approach to the metadata associated with image datasets. This standardization process delineates the structure, attributes, and terminology of such metadata. We are presenting this evolving standard, which has been under continuous development since 2022. Alongside it, we are introducing a suite of applications and integrations that we and the community developed to facilitate comprehensive lifecycle management of iFDOs. Our findings indicate that the adoption of iFDOs not only augments the quality of image data publication but also refines the overall workflow involved in the creation and publication of datasets.

Keywords: FAIR Digital Object, Marine Imaging

1. Background

Images, by their very nature, serve as a rich source of information; however, to accurately extract and comprehend their contents, it is imperative to incorporate additional metadata regarding the context in which these images were acquired. Certain scientific disciplines exhibit relatively uniform imaging contexts, such as in optical astronomy. Conversely, marine imaging represents a highly heterogeneous domain, encompassing subjects ranging from plankton to ice sheets. The imaging platforms employed vary from deep-sea remotely operated vehicles to unmanned aerial vehicles and aircraft, and the associated scientific disciplines span from microbiology to glaciology. Harmonizing these diverse contexts into a unified metadata standard not only facilitates the FAIR (Findable, Accessible, Interoperable, and Reusable) publication of numerous image and video datasets but also enables the utilization of marine image data across various scientific domains. By specifying image content in such a manner, machine learning initiatives are empowered, while also enhancing the findability of datasets with standardized image annotations.

2. iFDO Standard

The FAIR Digital Objects for Images (iFDO) standard is designed to harmonize marine imaging metadata by providing a FAIR Digital Object (FDO) profile for image and video datasets **ifdo**. Technically, an iFDO consists of a plain text file, which inherently does not contain the image data itself but rather references the dataset and the individual images and videos through the use of persistent identifiers, such as handles. To facilitate the referencing of individual images and videos to an iFDO, even when they are downloaded or dissociated from their dataset, the iFDO standard mandates the inclusion of a universally unique identifier (UUID) within both the object's metadata header and the iFDO itself.

The structure and attributes of the iFDO are specified using a JSON schema. Structurally, it is separated into two main components: the "image-set-header," which encompasses all metadata relevant to the entire dataset, and the "image-set-items" section, which provides detailed information specific to each individual item. The mandatory attributes encompass references to the dataset and image items previously mentioned, alongside basic contextual information, including the device utilized for producing the image dataset and the contributors involved in its creation, as well as geospatial and legal information. Additionally, many of these attributes accommodate the referencing of metadata repositories and FDOs, such as ORCID, through the utilization of persistent identifiers (PIDs). The recommended attributes furnish supplementary context regarding the methodology of image item capture, and they describe the content of the images using fundamental statistics, such as average color and annotations.

3. iFDO Life Cycle and Ecosystem

The development of metadata documents, while potentially perceived as an additional task for researchers, has been demonstrated to streamline the publication process when iFDO records are integrated early in the data lifecycle. It is essential, however, that the processes of record creation and dataset compilation are optimized for simplicity. The Marimba Python image dataset processing framework and pipeline runner, which have been developed by CSIRO¹ and MBARI², facilitate the transformation of raw data into data products with a high degree of automation, while also generating associated iFDO records **marimba**. We are in the process of developing the "iFDO Creator," an intuitive graphical user interface tool designed to simplify the creation and modification of iFDOs. This tool incorporates existing metadata repositories, enabling attribute auto-completion and allowing users to seamlessly integrate references to repositories, such as ORCID.

Following the establishment of the initial record, it can be employed by other software applications, such as the annotation tool BIIGLE **biigle**, which can further enhance the iFDO. Upon dataset completion, the iFDO record facilitates automated submissions to data repositories like PANGAEA **pangaea**. These repositories can deduce much of their required metadata from the iFDO, reducing the manual entry burden on researchers. Subsequently, the iFDO is published alongside the dataset, as exemplified in **boehringer2024sica**. From this stage, both the dataset and iFDO can be accessed by data viewers, such as the Marine Data Portal **Koppe'2015**, which enhances the discoverability of individual images through the metadata encapsulated in the iFDO record. This demonstrates that FDOs and iFDOs significantly enhance not only the value of published data but also the overall process of dataset creation and publication.

¹Commonwealth Scientific and Industrial Research Organisation

²Monterey Bay Aquarium Research Institute

Data availability statement

The example image dataset containing an iFDO is available at <https://doi.pangaea.de/10.1594/PANGAEA.971418> under Creative Commons Attribution 4.0 International license.

Underlying and related material

The current iFDO standard documentation can be found at <https://ifdo-schema.org/>.

Author contributions

Karsten Schimpf: Writing – original draft, Visualization, Software; Christopher Krämmer: Writing – review & editing, Software; Timm Schoening: Writing – review & editing, Project administration, Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Funding

This initiative is partly funded by the DataHub initiative supported by the information infrastructure for integrated Earth system research 2021-2025 of the Helmholtz Association Earth and Environment research field as well as by the Helmholtz Strategic Infrastructure MUSE - Marine Environmental Robotics and Sensor Technologies for sustainable Research and Management of Coastal, Ocean, and Polar Regions, 2023-2029.

Acknowledgements

The AI tools "languagetool.org", "deepL Write" and "Writefull for Overleaf" were used for spell, grammar, and style checking of this document.